l^q-regularization of the Kalman Filter for exogenous outlier removal: application to hedge funds analysis

Emmanuelle Jay*§, Patrick Duvaut*†§, Serge Darolles‡§ and Christian Gouriéroux‡§*

*QAMLAB, Palais Brongniart, 28 Place de la Bourse, 75002 Paris, France, Email: emmanuelle.jay@qamlab.com
†ETIS/CNRS UMR 8051, 6 av. du Ponceau, 95014 Cergy-Pontoise Cedex, France, Email: patrick.duvaut@ensea.fr
‡CREST, 15 Bd Gabriel Péri, 92245 Malakoff Cedex, France, Email: serge.darolles@ensae.fr
§QUANTVALLEY, Palais Brongniart, 28 Place de la Bourse, 75002 Paris, France

Abstract—**This paper presents a simple and efficient exogenous outlier *detection & estimation* algorithm introduced in a regularized version of the Kalman Filter (KF). Exogenous outliers that may occur in the observations are considered as an additional stochastic impulse process in the KF observation equation that requires a regularization of the innovation in the KF recursive equations. Regularizing with a l^1- or l^2-norm needs to determine the value of the regularization parameter. Since the KF innovation error is assumed to be Gaussian we propose to first detect the possible occurrence of an exogenous impulsive spike and then to estimate its amplitude using an adapted value of the regularization parameter. The algorithm is first validated on synthetic data and then applied to a concrete financial case that deals with the analysis of hedge fund returns. The proposed algorithm can detect anomalies frequently observed in hedge returns such as illiquidity issues.**

I. Introduction

Exogenous outlier removal is one of the most common problems encountered in processing real data. Outliers may come from unknown sensor failures, abnormal measurements or even intentional jamming. In the framework of tracking or estimating unobserved processes, usually solved with KF or one of its extensions, such exogenous outliers are referred to as *additive outliers* or *AO*'s [1], [2] and occur in the observation. Endogenous outliers that may affect the innovations of the state equation are called *innovative outliers* or *IO*'s and are out of the scope of this paper. Recursive least squares can serve to deal with AO's applying a Huber weight function [3] to the innovation error in the KF recursive equation. One can also penalize the Kalman filter by introducing a l^1- or l^2- regularization term. Regularization usually offers good results but requires the determination of the regularization parameter introduced to limit the amplitude of the solution. The search for the optimal regularization parameter has no ideal solution and each existing method (e.g. cross-validation, generalized cross-validation, maximum a posteriori estimation, Bernoulli-Gauss assumption or MCMC methods) is fit to the problem under study. Regularization is widely used in statistical signal processing (e.g. in basis pursuit, compressed sensing, signal recovery, wavelet thresholding), in statistics with ridge regressions and LASSO algorithm or fused LASSO [4] and is also very useful in decoding linear codes or geophysics problems. In finance, l^1-regularization is also leveraged in the context of portfolio optimization [5], [6], [7].

The paper is organized as follows: section 3 presents the robust Kalman Filter (RKF) as described in [8]. The family of *detection & estimation* algorithms we introduce to handle exogenous outliers is presented in section 4 and section 5 is devoted to the application of our algorithm to detecting irregularities in hedge funds returns. The sixth section concludes.

II. Notation

Uppercase and lowercase bold letters denote matrices and column-vectors, respectively. $(.)'$ stands for transpose operator, $\mathbf{0}_J$ for a $J \times J$ null matrix, $\mathbf{I}_J$ for a $J \times J$ identity matrix, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\mathbb{E}(.)$ for the mathematical expectation, $Var(.)$ for the variance, $Cov(.)$ for the covariance, $||.||_1$ for the l^1-norm, and $||.||_2$ for the l^2-norm.

III. Robust Kalman Filter (RKF)

The Robust Kalman Filter (RKF) setup is as follows [8]:

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \tag{1}$$

$$r_t = \mathbf{g}_t' \boldsymbol{\theta}_t + v_t + \epsilon_t. \tag{2}$$

$\boldsymbol{\theta}_t$ is the unknown m-dimensional state vector to be estimated at each time step t, $\boldsymbol{\omega}_t$ - the state noise - is $\mathcal{N}(\mathbf{0}_m, \sigma_w^2 \mathbf{I}_m)$, r_t is the observed value at time t, $\mathbf{g}_t$ is a m-dimensional vector of observed variables, ϵ_t - the observation noise - is $\mathcal{N}(0, \sigma_\epsilon^2)$, and v_t is a stochastic impulse process, e.g. a Bernoulli-Gaussian process, suited to model unknown sensor failures, measurement outliers or even intentional jamming. Parameters σ_w and σ_ϵ are supposed to be known and the observed variable r_t is scalar.

As with the standard KF (where $v_t = 0$), the objective of the problem stated in (1) and (2) is to derive recursive estimate $\hat{\boldsymbol{\theta}}_{t/t}$ of the unobservable parameter $\boldsymbol{\theta}_t$, given the measurement r_t and the observed variables $\mathbf{g}_t$ up to time t, based on the minimization of the mean-square error (MSE) $\mathbb{E}(|\boldsymbol{\theta}_t - \hat{\boldsymbol{\theta}}_{t/t}|^2)$. In the presence of the non-Gaussian outlier v_t, the optimum recursive estimate cannot be explicitly derived and moreover may be nonlinear. An elegant alternative consists in regularizing the KF quadratic cost function by adding a l^q-penalty (l^1 in RKF). The RKF algorithm goes through the

following steps. The prediction step is the same as for the standard KF and is given by:

KF Prediction

$$\hat{\boldsymbol{\theta}}_{t/t-1} = \hat{\boldsymbol{\theta}}_{t-1/t-1}, \quad (3)$$

$$\mathbf{S}_{t/t-1} = \mathbf{S}_{t-1/t-1} + \sigma_w^2\,\mathbf{I}_m, \quad (4)$$

where $\mathbf{S}_{t/t} = Cov(\boldsymbol{\theta}_t - \hat{\boldsymbol{\theta}}_{t/t})$ is the error covariance matrix of the estimates. To handle the additional term v_t in (2), the recursive equations of the correction step are slightly modified. Given the innovation or measurement error:

$$\tilde{r}_t = r_t - \mathbf{g}_t'\,\hat{\boldsymbol{\theta}}_{t/t-1}, \quad (5)$$

and the *optimal* Kalman gain:

$$\mathbf{z}_t = \mathbf{S}_{t/t-1}\,\mathbf{g}_t\,(\mathbf{g}_t'\,\mathbf{S}_{t/t-1}\,\mathbf{g}_t + \sigma_\epsilon^2)^{-1}, \quad (6)$$

then, the correction equations are as follows:

Correction

$$\mathbf{S}_{t/t} = (\mathbf{I}_m - \mathbf{z}_t\,\mathbf{g}_t')\,\mathbf{S}_{t/t-1}, \quad (7)$$

$$\hat{\boldsymbol{\theta}}_{t/t} = \hat{\boldsymbol{\theta}}_{t/t-1} + \mathbf{z}_t\,(\tilde{r}_t - \hat{v}_t). \quad (8)$$

In (8), the estimated value $\hat{v}_t$ is the solution of:

$$\underset{v_t}{\arg\min}\, C(v_t,\delta) = \min_{v_t}(\tilde{r}_t - v_t)'\, q_t\,(\tilde{r}_t - v_t) + \delta\,||v_t||_1, \quad (9)$$

with $q_t = (1 - \mathbf{g}_t'\,\mathbf{z}_t)'\,\sigma_{\epsilon_i}^{-2}\,(1 - \mathbf{g}_t'\,\mathbf{z}_t) + \mathbf{z}_t'\,\mathbf{S}_{t/t-1}^{-1}\,\mathbf{z}_t > 0$ and $\delta \geq 0$ being the regularization parameter.

Equation (9) is referred to as a *regularized regression* and is equivalent to the LASSO setup [4], [9]. Given an appropriate value of δ, minimizing $C(v_t,\delta)$ is equivalent to jointly detecting and estimating v_t. The l^1-norm will shrink the value of v_t to zero for small values of δ and consider a nonzero value of v_t with a high value of δ. Jointly detecting and estimating $\hat{v}_t$ relies therefore on the good choice of δ which has to be tuned such that the sparsity of v_t coincides with the sparsity seen through the observations. If the value of δ is inappropriate then many false alarms are generated. In [4], the author observes that the search for δ can be restricted to the range $[0, 2\,q_t\,|\tilde{r}_t|]$. In practice, the upper bound seems to be very generous [10], but it can be adaptive and specific to the observed variable.

IV. REGULARIZED KALMAN FILTER (rgKF)

Hereafter, we propose a family of algorithms called the *regularized Kalman Filter* or rgKF(OD, E-l^q) that includes an *Outlier Detection* step (OD) followed by a l^q ($q > 0$) *Estimation* step (E-l^q). OD stands for the name of the detection method, e.g. if OD=NG then a Non-Gaussian deviation test is applied. In the following, we consider only $q = 1, 2$. The RKF in [8], which consists of a single and joint detection/estimation step, falls into the l^1-regularization approach and is equivalent to rgKF(l^1, E-l^1). TABLE I sums up these algorithms.

What we call the *regularized Kalman Filter* or rgKF is a simple and efficient algorithm. Based on the fact that v_t is an additive outlier in the observations, it is obvious that

	OD	**E-l^q**	δ
RKF		l^1	fixed
rgKF(NG, E-l^1)	NG	l^1	$\delta \in [0, 2\,q_t\,|\tilde{r}_t|]$
rgKF(NG, E-l^2)	NG	l^2	fixed

the innovation $\tilde{r}_t$ in (5) should contain v_t if v_t occurs at time t. Moreover, the observation error ϵ_t is assumed to be a Gaussian random variable with a fixed variance σ_ϵ^2. If v_t occurs at time t, then $\tilde{r}_t$ should not lie in the range space of the Gaussian density. Our rgKF algorithm with exogenous outlier *detection & estimation* can be summed-up in the four following steps:

---

rgKF(NG, E-l^q), $q = 1, 2$

---

1) KF Prediction
 a) Computation of $\hat{\boldsymbol{\theta}}_{t/t-1}$ (3) and $\mathbf{S}_{t/t-1}$ (4)
 b) Computation of innovation (5) and Kalman Gain (6),

2) Outlier Detection
 a) Prediction step of the mean and standard deviation of $\tilde{r}_t$ (for $t > 2$, with initial values $\hat{\mu}_{2/2} = \tilde{r}_2$ and $\hat{\sigma}_{2/2}^2 = 0$):

$$\hat{\mu}_{t/t-1} = \frac{t-1}{t}\hat{\mu}_{t-1/t-1} + \frac{1}{t}\tilde{r}_t, \quad (10)$$

$$\hat{\sigma}_{t/t-1}^2 = \frac{t-2}{t-1}\hat{\sigma}_{t-1/t-1}^2 + t\,(\hat{\mu}_{t/t-1} - \hat{\mu}_{t-1/t-1})^2. \quad (11)$$

 b) Non-Gaussian test: if $|\tilde{r}_t - \hat{\mu}_{t/t-1}| > 3\,\hat{\sigma}_{t/t-1}$ then an outlier is detected.

3) Outlier Estimation
 • if **2b)** is verified:
 E-l^1 case: set $\delta_t = 10^{-2}$ and estimate $\hat{v}_t$ using (9).

 E-l^2 case: if the regularization term in (9) is $\delta\,||v_t||_2^2$ instead of $\delta\,||v_t||_1$ then estimate $\hat{v}_{t,l^2}$ is as follows:

$$\hat{v}_{t,l^2} = \frac{\tilde{r}_t\,q_t}{\delta_{,l^2} + q_t}. \quad (12)$$

Note: by definition and following **2b)**, *we should have* $|\tilde{r}_t - \hat{v}_{t,l^2}| < 3\,\sigma_\epsilon$, *so that* δ_{t,l^2} *should lie in the range* $\left[0, \dfrac{3\,\sigma_\epsilon\,q_t}{|\tilde{r}_t| - 3\,\sigma_\epsilon}\right]$

 • otherwise (if no detection) $\hat{v}_t = 0$.

4) rgKF Correction
 - The covariance matrix is updated using (7),
 - The state is updated according to (8),
 - The mean of the innovation is updated as follows:

$$\hat{\mu}_{t/t} = \hat{\mu}_{t/t-1} - \frac{1}{t}\hat{v}_t, \quad (13)$$

 - The variance of the innovation becomes:

$$\hat{\sigma}_{t/t}^2 = \frac{t-2}{t-1}\hat{\sigma}_{t-1/t-1}^2 + t\,(\hat{\mu}_{t/t} - \hat{\mu}_{t-1/t-1})^2. \quad (14)$$

V. APPLICATION TO DETECTING IRREGULARITIES IN HEDGE FUNDS RETURNS

A. Definition and description of the problem

In the financial asset literature, (2) with $v_t = 0$ is known as a *multi-factor model* (see [11] for more details) and is statistically interpreted as a linear regression with stochastic parameters. The returns r_t of a risky asset are linked to some factors $\mathbf{f}_t$ (in (2), $\mathbf{g}_t = \begin{bmatrix} 1 & \mathbf{f}_t' \end{bmatrix}'$ to include a constant term in the regression and $\boldsymbol{\theta}_t = \begin{bmatrix} \alpha_t & \boldsymbol{\beta}_t' \end{bmatrix}'$) which may be micro- or macro-economic data, fundamental values, statistical factors or specific portfolios. These factors represent the various sources of risk present in the market to which an investor is exposed. The risks coming from the factors exposures $\boldsymbol{\beta}_t$ are called the *systematic risks*: they are common to all of the assets whose returns are expressed using (2) and cannot be lowered by diversification. Oppositely, the residual terms in (2) are totally specific to the assets and are related to the so-called *idiosyncratic risks*. They represent the part of the returns which cannot be explained by market factors but whose risks may be reduced with diversification. TABLE II gives an overview of the asset returns and risk break down.

TABLE II
BREAK DOWN OF ASSET RETURNS AND RISKS INTO *systematic* AND *idiosyncratic* COMPONENTS

	Systematic Term	Idiosyncratic Term
Asset returns	Market returns	Asset-specific returns
r_t	$\mathbf{f}_t' \boldsymbol{\beta}_t$	$\alpha_t + v_t + \epsilon_t$
Risks	Market Risks[(a)]	Idiosyncratic risks
$Var(r_t)$	$\boldsymbol{\beta}_t' \boldsymbol{\Sigma}_f \boldsymbol{\beta}_t$	$\sigma_v^2 + \sigma_\epsilon^2$

[(a)] $\boldsymbol{\Sigma}_f$ *is the covariance matrix of the factors.*

The goal of our example is to explain the importance of removing outliers using *rg*KF algorithms. A multi-factor decomposition of the returns helps us to understand which part of the returns comes from the market and which one is specific. If outliers appear in the observations and if the model does not take them into account, then market exposures estimates can be totally spurious. *rg*KF algorithms avoid the contamination of these estimates by simply removing outliers. In practice, portfolio managers can use multi-factor approaches to hedge one specific or all market risks. Bad estimates can lead to take some inappropriate decisions concerning the optimal hedging portfolio. The tests are first conducted on synthetic data and then applied on real data.

B. Impact of outliers on synthetic returns

Synthetic returns are built with constant parameters $\alpha = 0.02$, $\beta_1 = -0.15$ and $\beta_2 = 0.23$, and two of the eight risk factors of Fung & Hsieh (F&H) [12], [13], the Standard and Poor's 500 Index (S&P500, $\mathbf{f}_1$) and the Emerging Market index (MXEF, $\mathbf{f}_2$). A Gaussian noise is added to the resulting time series. A spiky version of these synthetic returns is also constructed (adding occurences of v_t): we randomly chose three instants of arrival (being 30-Jun-2000, 31-Oct-2003 and 31-Dec-2007 on Fig. 1) on which spikes were randomly

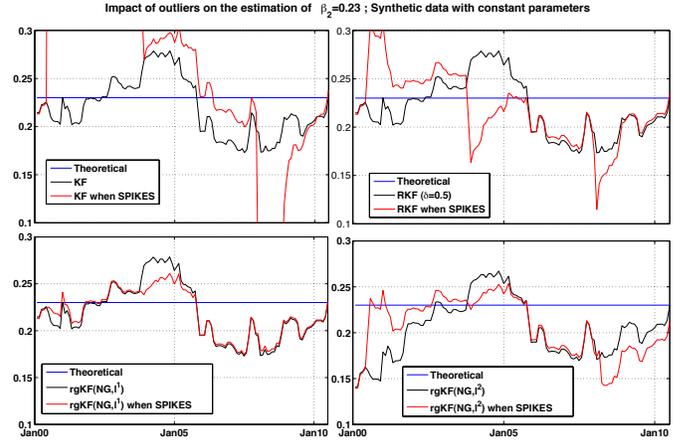generated as $\mathcal{N}(0, 0.5)$. Fig. 1 shows what happens on the

Fig. 1. Comparison of the impact of outliers present in the observations on the estimation of β_2 of the synthetic data between KF, RKF ($\delta = 0.5$), rgKF(NG,l^1) and rgKF(NG,l^2;$\delta = 0.5$).

estimated parameters of the model and compares four different estimation methods: KF, RKF with $\delta = 0.5$, rgKF(NG,l^1) and rgKF(NG,l^2) with $\delta = 0.5$. We observe that without any consideration for the possible occurrence of outliers in the observations (as for KF), parameter estimations become totally spurious and very far from the theoretical values. Inversely, rgKFs perform similarly with or without outliers and the results are also very close to the Kalman Filter results without outliers.

This procedure is run 1000 times in order to evaluate the MSEs of the parameter estimations for the different algorithms. The MSEs distributions are displayed on Fig. 2 as boxplots. For each simulation, we keep the maximum of the β's MSEs. As a benchmark method, we have compared the results to those obtained by using a 24-long rolling window ordinary least squares (OLS) regression to estimate the constant parameters. OLS is known to deliver the optimum solution in such a setting (regression with constant coefficient) and acts therefore as the lower bound for optimality. As expected, OLS gives the lowest estimation errors when no outlier is present, followed by KF in the same situation. When outliers are added, then rgKF(NG,l^1) gives a similar performance as in the no outlier case and tends to deliver the same performance as a KF with no outlier. RKF and rgKF(NG,l^2) performance depends on the choice of δ. Setting $\delta = 0.01$ for RKF yields the entire sets of observations to be considered as outliers.

C. Hedge funds returns

This section is devoted to applying rgKF(NG, l^1) on real hedge funds data to detect the potential occurrence of outliers. We conduct the analysis on returns of *Global Macro* hedge funds. Managers of such funds take directional positions in currencies, debt, equities, or commodities and may also elect to take relative positions combining long positions paired off against short positions. Hedge funds data comes from the
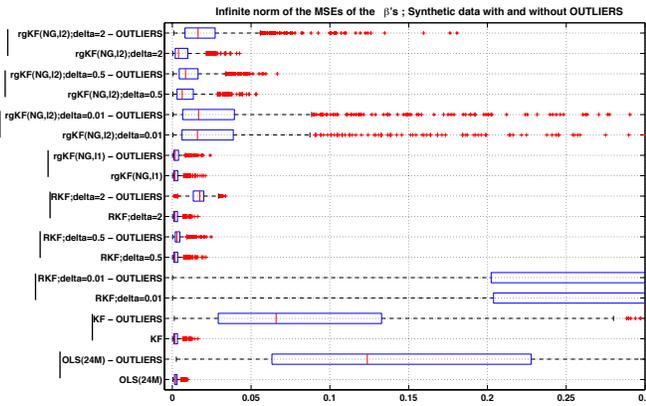
Fig. 2. Distribution of the 1000 maxima of the β's MSEs computed for each of the 1000 synthetic time series of returns of size 150, with or without additional outliers.

hedge fund research (HFR) database [14]. The returns are end-of-month returns, available from January 1998 to June 2010 (i.e. for 150 months).

We choose the eight risk factors of Fung & Hsieh (F&H) [15], selected and constructed by the authors [13] to explain trend-follower hedge funds. These factors are now widely used in the literature of hedge funds analysis. They are quoted on a monthly basis and stored in $\mathbf{f}_t$ at each end-of-month t.

Fig. 3 illustrates the capability of the rgKG(NG,l^1) to detect, estimate and filter some points in the observed returns time series which might be considered as hedge fund's specific events, outliers, irregularities or illiquidity issues. As it can be seen on the top graph of Fig. 3, when an outlier occurs, then the resulting KF estimations of β suddenly jump to four or five times their previous values whereas no special event occurs in the market.
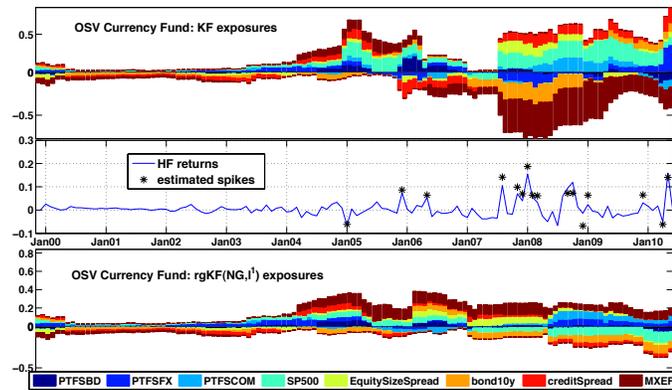

Fig. 3. Outliers detection and estimation for the *OSV Currency Fund* returns. Impact of outliers on the fund's exposures can be observed on the top graph of the figure: KF does not filter the rgKF(NG,l^1) estimated spikes and gives misled estimations of the market risks.

VI. CONCLUSION

This article has presented a new family of algorithms named rgKFs for *regularized KFs* which have been derived to *Detect*

and Estimate exogenous outliers that might occur in the observations equation of a standard Kalman Filter. Inspired from the Robust Kalman Filter (RKF) of [8] which makes use of a l^1-regularization step, we introduce a simple but efficient detection step in the recursive equations of the RKF. This solution is one way to solve the problem of adapting the value of the l^1-regularization parameter: when an outlier is detected in the innovation term of the KF, then the value of the regularization parameter is set to a value that will let the l^1-based optimization problem estimate the amplitude of the spike. We have also tested a l^2-based regularization term for which the solution is known in a closed form. The results obtained on synthetic data show that the best performance are achieved by the rgKF(NG,l^1) that handles the choice of the regularization parameter. In the context of analyzing hedge funds returns the rgKF(NG,l^1) could help risk managers to dynamically and accurately estimate the exposures of hedge funds for which they do not have any other information than the observed returns. Interpreting outliers as illiquidity issues, our algorithms would therefore be very useful to classify funds according to their "illiquidity quality" and to make accurate decisions from a regulation point of view.

REFERENCES

[1] A. J. Fox, "Outliers in time series," *Journal of the Royal Statistical Society, Series B*, vol. 34, pp. 350–363, 1972.

[2] P. Ruckdeschel, "Optimally robust kalman filtering," Berichte des Fraunhofer ITWM 185, Fraunhofer, ITWM, May 2010.

[3] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics, Theory and Methods*, John Wiley & Sons, 2006.

[4] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society*, vol. 58, no. Series B, pp. 267–288, 1996.

[5] R. Jagannathan and T. Ma, "Risk reduction in large portfolios: why imposing the wrong constraints helps," *Journal of Finance*, vol. 58, no. 4, pp. 1651–1684, 2003.

[6] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris, "Sparse and stable markowitz portfolios," September 2008, European Central Bank Working Paper No 936.

[7] V. DeMiguel, L. Garlappi, F. J. Nogales, and R. Uppal, "A generalized approach to portfolio optimization: improving performance by constraining portfolio norms," *Management Science*, vol. 55, no. 5, pp. 798–812, May 2009.

[8] J. Mattingley and S. Boyd, "Real-time convex optimization in signal processing," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 50–61, May 2010.

[9] M. R. Osborne, B. Presnell, and B. A. Turlach, "On the LASSO and its dual," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 319–337, June 2000.

[10] M. Schmidt, "Least squares optimization with L^1-norm regularization," Project Report CS542B, UBC, University of Alberta, Canada, December 2005.

[11] E. Jay, P. Duvaut, S. Darolles, and A. Chrétien, "Multi-factor models : examining the potential of signal processing techniques," *IEEE Signal Processing Magazine*, vol. 28, no. 5, September 2011.

[12] D. A. Hsieh, "Data library webpage: Hedge fund risk factors," http://faculty.fuqua.duke.edu/%7Edah7/HFRFData.htm.

[13] W. Fung and D. Hsieh, "The risk in hedge fund strategies, theory and evidence from trend followers," *Review of Financial Studies*, vol. 14, no. 2, pp. 313–341, 2001.

[14] Hedge Fund Research, "HFR databases," http://www.hedgefundresearch.com/.

[15] D. A. Hsieh, "Data library webpage: Hedge fund risk factors," http://faculty.fuqua.duke.edu/\%7Edah7/HFRFData.htm.